Multi-sensor fusion and enhancement using the Retinex image enhancement algorithm

Zia-ur Rahman[†], Daniel J. Jobson[‡], Glenn A. Woodell[‡], Glenn D. Hines[‡]

[†]College of William & Mary, Department of Computer Science, Williamsburg, VA 23187.
[‡]NASA Langley Research Center, Hampton, Virginia 23681.

ABSTRACT

A new approach to sensor fusion and enhancement is presented. The retinex image enhancement algorithm is used to jointly enhance and fuse data from long wave infrared, short wave infrared and visible wavelength sensors. This joint optimization results in fused data which contains more information than any of the individual data streams. This is especially true in turbid weather conditions, where the long wave infrared sensor would conventionally be the only source of usable information. However, the retinex algorithm can be used to pull out the details from the other data streams as well, resulting in greater overall information. The fusion uses the multiscale nature of the algorithm to both enhance and weight the contributions of the different data streams forming a single output data stream.

1. INTRODUCTION

The Multiscale Retinex (MSR) was developed as a general-purpose image enhancement algorithm that provided simultaneous dynamic range compression and color constancy. Aside from the obvious applications to the correction of color and grayscale digital images, this algorithm can also be applied to enhance infrared and near-infrared spectral bands. In addition, due to the inherent multiscale structure, it can be used to emphasize different features in different spectral bands, yet provide a composite image that contains more information than any individual spectral band. It is the combination of these properties that led us to investigate the MSR as a possible fusion engine. In this paper we will present the results of two different approaches for using the MSR as a fusion engine. The primary driving force for developing these fusion strategies was to provide an image that could be used to "see" in poor visibility conditions such as dusk, and fog.

The Retinex (retina+cortex) was developed as a model of the human color vision system by Edwin Land.[1–3] The MSR was developed by Jobson, et al[4, 5] to address the problems due to changing lighting and atmospheric conditions that were inherent to the process of acquiring images of Earth from space. The MSR provides the dynamic range compression, color constancy and sharpening that are required to alleviate these problems. It has been shown that applying the MSR to remotely-sensed images improves the overall classification accuracy.[6, 7] In addition, the MSR processed images allow a greater number of features to be detected than would be possible with un-enhanced data. Both of these results are also of importance in the context of sensor fusion because they provide more accurate registration points between the data streams that need to be fused.

The MSR belongs to the class of center-surround operators and can be succinctly expressed as

$$\mathcal{R}_i(x_1, x_2) = \sum_{k=1}^{K} \mathcal{W}_k\big(\log[\mathcal{I}_i(x_1, x_2)] - \log[\mathcal{I}_i(x_1, x_2) * \mathcal{S}_k(x_1, x_2)]\big), \quad i = 1, \ldots, \mathcal{N}, \tag{1}$$

where $\mathcal{R}_i$ and $\mathcal{I}_i$ are, respectively, the ith spectral band of the MSR output and the input, $\mathcal{W}_k$ and $\mathcal{S}_k$ are, respectively, the weight and the surround function associated with the kth scale, K is the number of scales, and $\mathcal{N}$ is the number of spectral bands. The kth surround function is given by

$$\mathcal{S}_k(x_1, x_2) = \alpha \exp[-(x_1^2 + y_1^2)/\sigma_k^2],$$

ZR: zrahman@cs.wm.edu; DJJ: d.j.jobson@larc.nasa.gov; GAW: g.a.woodell@larc.nasa.gov; GDH: g.d.hines@larc.nasa.gov

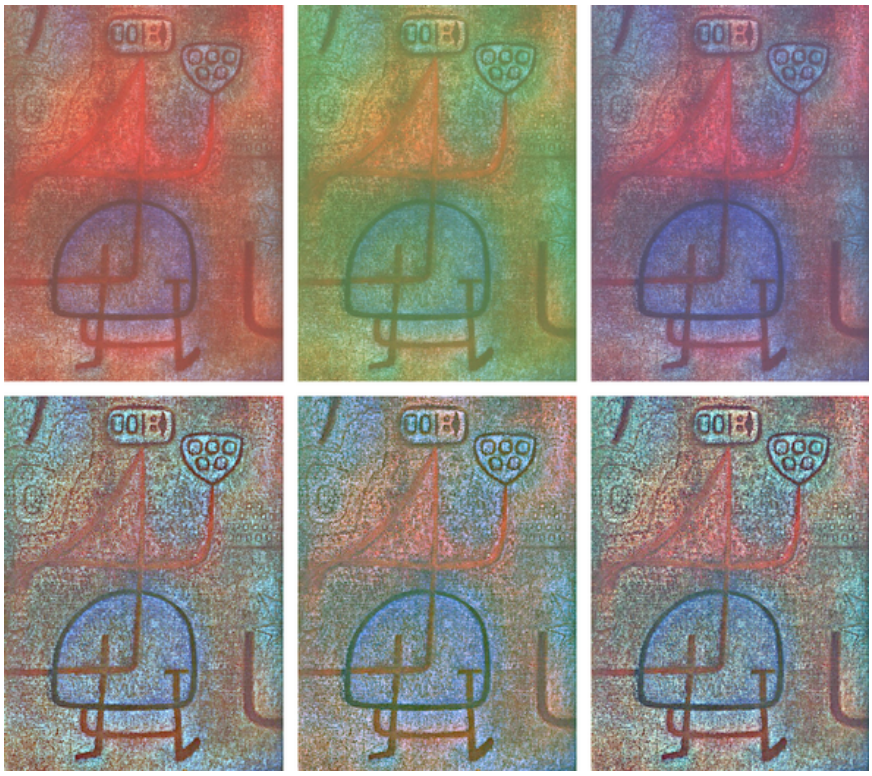

Figure 1: Example of color constancy and feature sharpening.

where σ_k is the scale factor that controls the width of the kth surround function, and $\alpha = \left(\sum_{x_1,x_2} \mathcal{S}_k(x_1,x_2)\right)^{-1}$ is the normalization factor. The color constancy of the MSR is a direct result of the center-surround nature of the algorithm. Equation 1 can be rewritten as:

$$\mathcal{R}_i(x_1,x_2) = \sum_{k=1}^{K} \mathcal{W}_k \left(\frac{\log[\mathcal{I}_i(x_1,x_2)]}{\log[\mathcal{I}_i(x_1,x_2) * \mathcal{S}_k(x_1,x_2)]} \right), \quad i = 1,\ldots,\mathcal{N}. \tag{2}$$

The intensity values in each spectral band at location (x_1,x_2) can be written as the product of two components: $\rho(x_1,x_2)$, the reflectance component, which represents the light reflected or emitted from all the objects in the scene, and $\iota(x_1,x_2)$ which represents the illumination component. That is,

$$\mathcal{I}_i(x_1,x_2) = \iota_i(x_1,x_2)\rho(x_1,x_2).$$

Since the illumination component varies very slowly across the scene, $\iota(x_1,x_2) \approx \mathcal{I}_o$, and $\mathcal{I}(x_1,x_2) \approx \mathcal{I}_o\rho(x_1,x_2)$. Thus,

$$\mathcal{R}(x_1,x_2) = \log\left(\frac{\mathcal{I}_o\rho(x_1,x_2)}{\mathcal{I}_o\rho(x_1,x_2) * \mathcal{S}(x_1,x_2)}\right) = \log\left(\frac{\rho(x_1,x_2)}{\rho(x_1,x_2) * \mathcal{S}(x_1,x_2)}\right),$$

which is independent of the illuminant. An example of this is shown in Figure 1 where the top row shows an image acquired under three different illumination conditions* and the bottom row shows the MSR outputs. Aside from the color constancy, the features in each of the processed images are much sharper, providing the common registration points that are necessary for image fusion.

*Unfortunately, the illumination effects are not readily visible in the grayscale rendition of the color images.

2. SENSOR FUSION

The particular problem that we are trying to address is to fuse data from three different imaging sources: long-wave infrared (LWIR), short-wave infrared (SWIR), and visible color (RGB). In addition to imaging at a different wavelength, these sensors also differ in resolution, and field of view (FOV). Thus, the first processing step is to perform a multi-modal registration of the images from the three different sensors. Table 1 shows the characteristics of the various sensors.[8] Image registration can generally be defined as a mapping between two

	LWIR	**SWIR**	**CCD**
Pixel Resolution (nominal)	$320H \times 240V$	$320H \times 240V$	$542H \times 497V$ (RGB)
Optics FOV	$39°H \times 29°V$	$34°H \times 25°V$	$34°H \times 25°V$
Detector readout frame rate	60Hz	60Hz (typical)	30Hz (interlaced)

Table 1: Selected sensor characteristics

or more images both spatially and with respect to intensity. Mathematically $I_2 = g(I_1(f(x, y)))$ where I_1 and I_2 are two-dimensional images (indexed by x and y), f is a transformation of spatial coordinates and g is a one-dimensional intensity or radiometric transformation.[9]

For the spatial transformation our primary complication is that there are no fiducial markers within the images. The cameras are, however, bore-sighted so they all have a common center of alignment. Thus we simply calculate an affine transformation performing the operations of scaling, translation and rotation globally on the images matching them to a common grid. A radiometric transformation could also be performed at this stage since grayscale characteristics, in particular polarity reversal between visible and IR images, may differ locally. Instead, as will be discussed in the next section, we utilize the MSR transformation. Further details regarding the rectification process are not discussed in this paper. We will assume, however, that appropriate rectification has been performed on each of the data streams with the resultant registration of features and correction of FOV and pixel resolution.

2.1. The first approach

Our first approach is relatively straightforward. Using the MSR as the enhancement engine, each of the data streams is individually enhanced to bring out the desired features. The MSR has previously been optimized for the RGB images but not for the LWIR, and SWIR images. So the first task was to analyze the data and determine the optimal parameters that were needed to achieve the desired enhancement in the LWIR and SWIR. This was done experimentally by applying the MSR given in Equation 1 to the LWIR and SWIR images, and observing the output images. The optimization, in this case, is simply the observer's judgment that a given image has more "useful" information than any others observed.

Once the optimized, MSR enhanced images were obtained, the RGB image was converted to grayscale (GS) by using the standard RGB to GS transform:

$$GS = 0.299R + 0.587G + 0.114B \tag{3}$$

where R, G, and B are the red, green, and blue color components of the RGB image respectively. Consistent with the contrast sensitivity of the human visual system,[10] the green channel is weighted most heavily when forming the GS image. After the conversion from RGB to GS, we have three remaining bands of data—GS, LWIR, SWIR—that we need to fuse. A quick way to do this is to use Equation 3 and use the three data bands in place of the R, G, and B components. The obvious question that arises is which data band should be substituted for which color component. Using the contrast sensitivity analogy, the data band which replaces the G component will have the most weight, so the fused image will be most sensitive to changes in this component. It is difficult to determine beforehand, which of the three bands contains the most relevant information and hence should be weighted most heavily. However, since the thrust for the sensor fusion is to provide better visibility in poor visibility conditions, we cannot, in general, rely on the GS band to provide the primary information. This task will typically fall to the LWIR band, especially under foggy conditions. The GS band does have the potential to

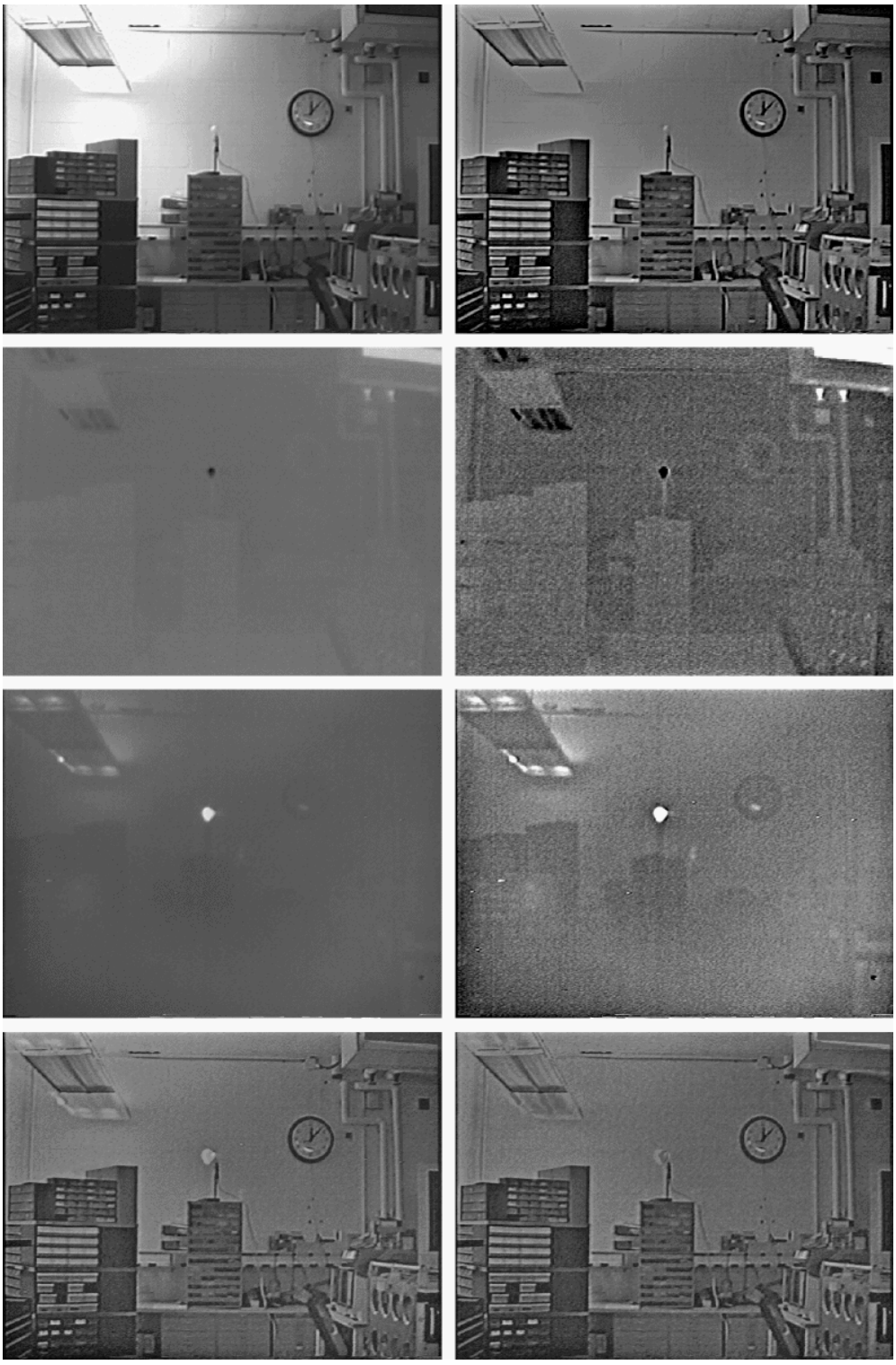

Figure 2. Original (left column) and MSR enhanced images (right column) are shown for RGB (first row), LWIR (second row), and SWIR (third row) images. The bottom row shows the composite images obtained by giving the strongest weight to the GS band: the left column shows the image obtained when the LWIR band is given the next heaviest weight and the right column shows the image when the SWIR band is given the next heaviest weight.

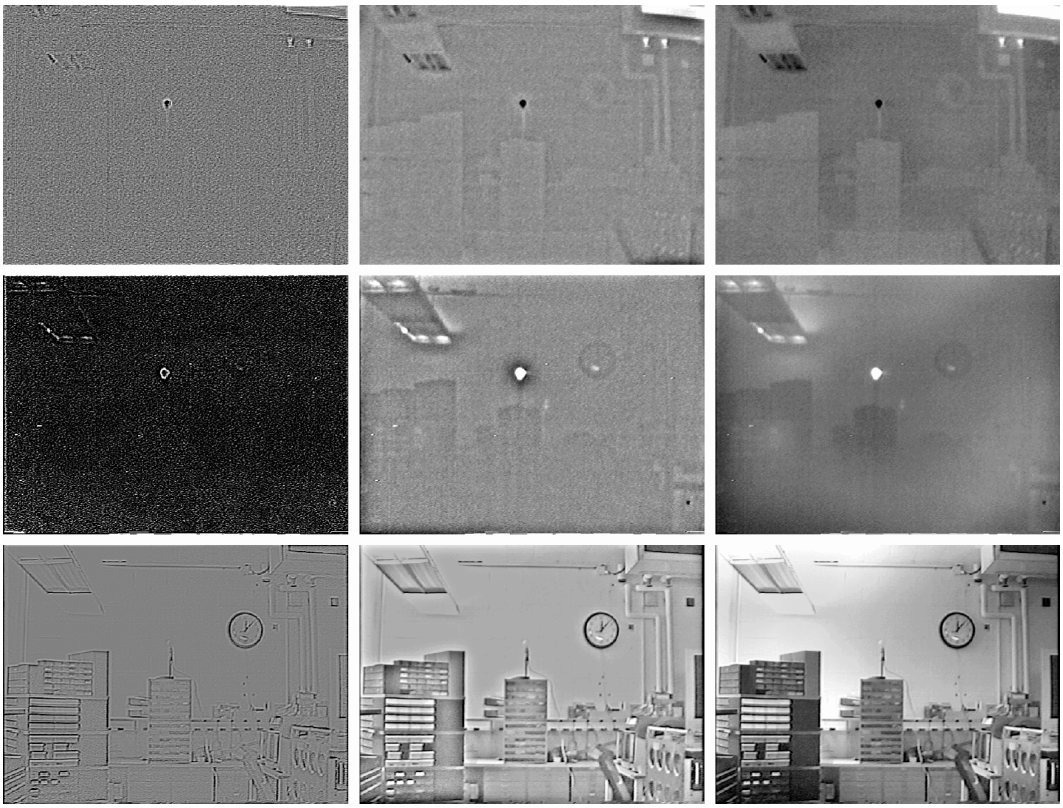

Figure 3. First column: MSR with $\sigma = 5$; middle column: MSR with $\sigma = 20$; last column: MSR with $\sigma = 240$. Top row: LWIR; middle row: SWIR; bottom row: GS. The narrow surround, $\sigma = 5$, acts as a high-pass filter, capturing all the fine details in the image but at a severe loss of tonal information. The wide surround, $\sigma = 240$, captures all the fine tonal information but at the cost of dynamic range. The medium surround, $\sigma = 20$, captures both dynamic range and tonal information.

provide the relevant information under low light conditions, such as dusk. The fused image, I_F, is then obtained by:

$$I_F = 0.299\text{SWIR} + 0.587\text{LWIR} + 0.114\text{GS}, \tag{4}$$

though other combinations are also possible.

Figure 2 shows the original and MSR enhanced LWIR, SWIR, and GS images, as well as the fused data I_F for images that were taken in a laboratory with lights dimmed to simulate lighting conditions similar to those at dusk. In this situation, the original GS has the most information but suffers from lack of information in certain regions in the image such as the top-left where the light from the fluorescent source saturates the image. Both the SWIR, and the LWIR images have information in this region. The MSR enhanced data provides more visual information in each of the bands—see the first three rows in the left column in Figure 2. However, the data from the SWIR and LWIR sensors is difficult to interpret on its own. The bottom row of Figure 2 shows two fused images obtained by:

$$I_{F_1} = 0.299\text{LWIR} + 0.587\text{GS} + 0.114\text{SWIR} \tag{5}$$

$$I_{F_2} = 0.299\text{SWIR} + 0.587\text{GS} + 0.114\text{LWIR} \tag{6}$$

where I_{F_1} is the image shown in the bottom-left of Figure 2 and I_{F_2} is show in bottom-right. Note in particular the additional information in the fused image around the two light sources: this information was missing in the GS image. Both of the fused images provide more information than any of the three data bands can individually.

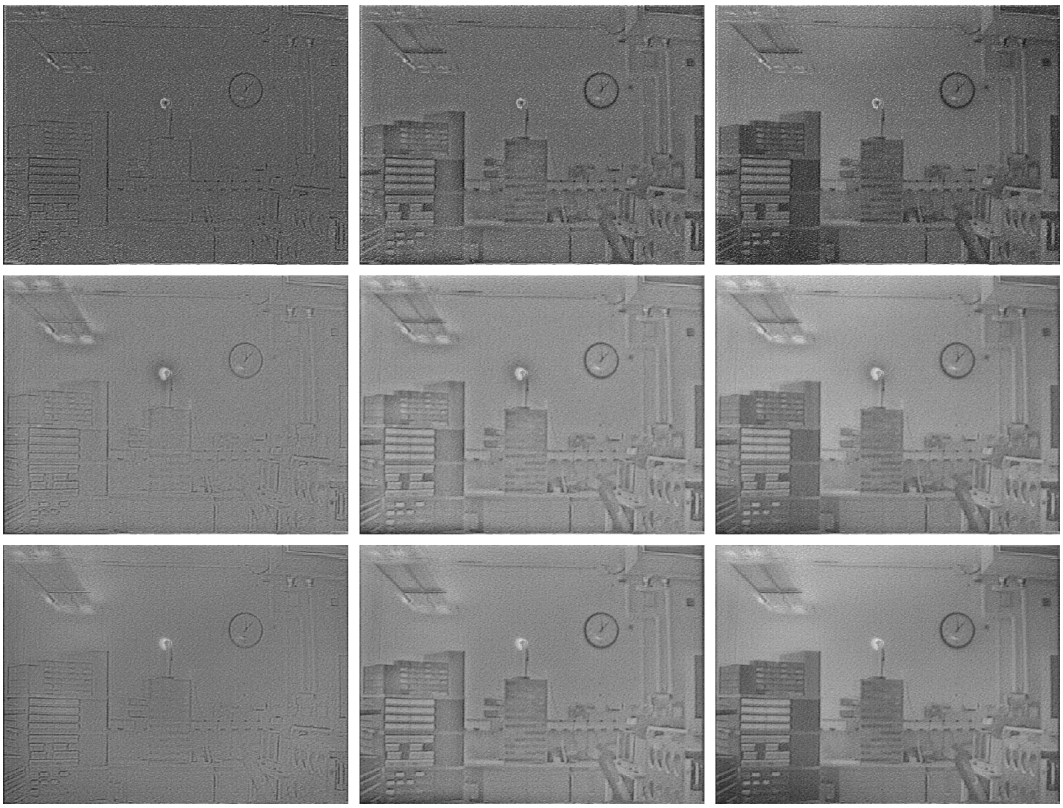

Figure 4. The fused images: $(\sigma_L, \sigma_S, \sigma_G)$ represents the scale values with which the LWIR, SWIR, and GS bands are processed for the SSR. Top row: $(5, 5, 5)$, $(5, 5, 20)$, $(5, 5, 240)$. Middle row: $(5, 20, 5)$, $(5, 20, 20)$, $(5, 20, 240)$. Bottom row: $(5, 240, 5)$, $(5, 240, 20)$, $(5, 240, 240)$.

.

Additionally, the compactness of the fused data, one image instead of three, also makes it more practical for the pilot in terms of ease of viewing and interpretation.

2.2. The second approach

Though the results from the first approach are quite encouraging, we wanted to test a second approach that uses the MSR in a different way. The first approach primarily relied on the dynamic range compression capabilities of the MSR to pull out features that were "difficult" to see in the existing lighting conditions, and then used Equations 5, and 6 to weight the contributions of the different data bands. In our second approach, we use a different mechanism. The MSR as shown in Equation 1 can be rewritten as

$$\mathcal{R}_i(x_1, x_2) = \sum_{k=1}^{K} \mathrm{SSR}_k(\mathcal{I}_i)(x_1, x_2) \quad i = 1, \ldots, \mathcal{N}, \tag{7}$$

where $\mathrm{SSR}_k(\mathcal{I}_i)$ is the single scale retinex (SSR) with the kth scale factor acting on the ith input band $\mathcal{I}_i$. As stated above, the magnitude of the kth scale factor, σ_k, controls the width of the surround function. This, in turn, determines the type and amount of "enhancement" that takes place: a large σ_k retains more color information at the cost of spatial detail, and a small σ_k enhances the spatial detail at the cost of color information. The MSR approach was developed to balance this tradeoff between color information and spatial detail by merging representations that brought out detail and color. Hence small details in the scene can be obtained by using using the MSR with $\mathcal{N} = 1$ and $K = 1$, and a small value of σ_k. This tradeoff between color information and spatial detail is shown in Figure 3.

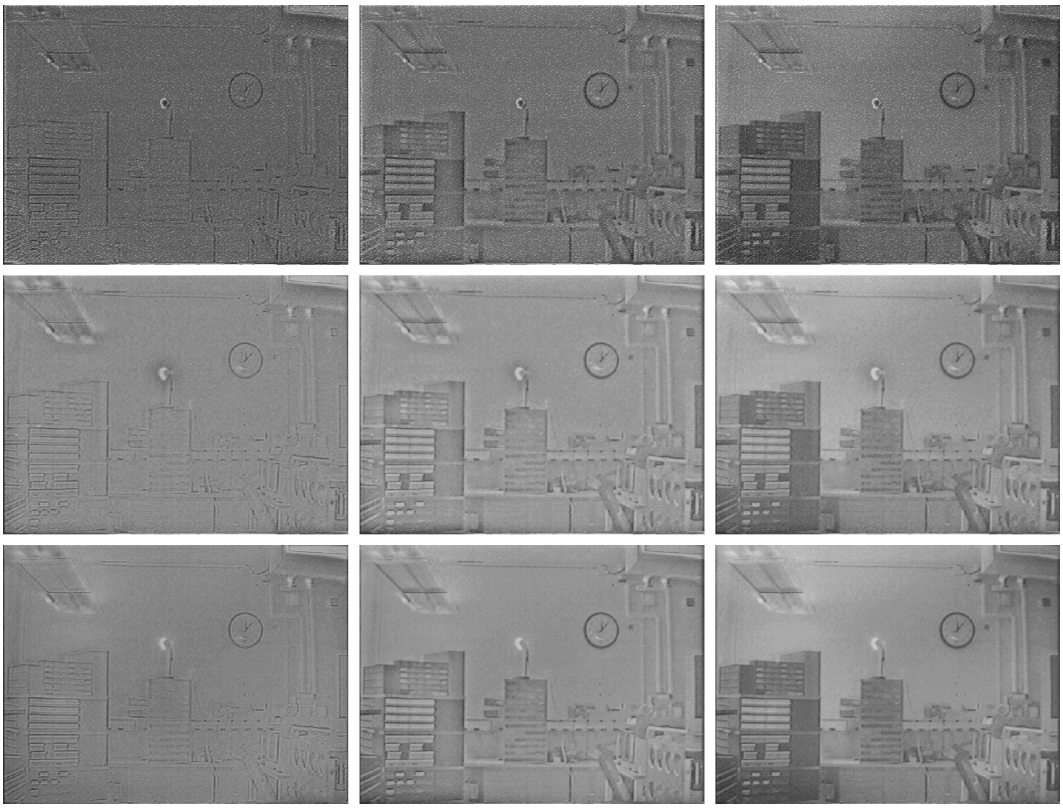

Figure 5. Top row: $(20, 5, 5)$, $(20, 5, 20)$, $(20, 5, 240)$. Middle row: $(20, 20, 5)$, $(20, 20, 20)$, $(20, 20, 240)$. Bottom row: $(20, 240, 5)$, $(20, 240, 20)$, $(20, 240, 240)$.

Using this formulation, we used the SSR to pull out different details from the different data bands before fusing the data together. The fused images can be obtained by averaging the results of any three of these nine renditions that are shown in Figure 3, with the condition that one image from each of the SWIR, LWIR, and GS bands must be used. So, for example, the fused image can be obtained by

$$I_F = \big(\mathrm{SSR}_1(\mathrm{LWIR}) + \mathrm{SSR}_2(\mathrm{SWIR}) + \mathrm{SSR}_3(\mathrm{GS})\big)/3 \tag{8}$$

where SSR_1 was generated with $\sigma_1 = 5$, SSR_2 was generated with $\sigma_2 = 20$, and SSR_3 was generated with $\sigma_1 = 80$. However, with the three bands each processed at three different scales, there is a total of 27 combinations that are possible: Figures 4–6 show all of these various combinations. As is readily evident, some of these combinations are not very useful when used for visual interpretation. However, the fine details in the combinations using the small scales maybe ideal for projection on a heads up display (HUD). The best visual result is obtained when $\sigma_L = 240, \sigma_S = 20$, and $\sigma_G = 240$ (Figure 6). This makes intuitive sense as well because in this case, the GS scale image provides the tonal information for the rendition, and the LWIR, and SWIR images provide the complementary features that are not visible in the GS band. A closer look at Figure 3 shows the complementary features in the three data bands. The amount of visual information that can be obtained from the fused image, as in the first experiment, is considerably more than the amount of visual information available in the individual data bands or in their SSR enhanced versions. This approach has the additional merit in that it allows extraction of fine features from the various data bands, hence making the task of seeing objects, large *and* small, in poor visibility conditions considerably easier. Interestingly, the best result in the second approach is obtained when the second approach is, in essence, a single scale version of the first approach.

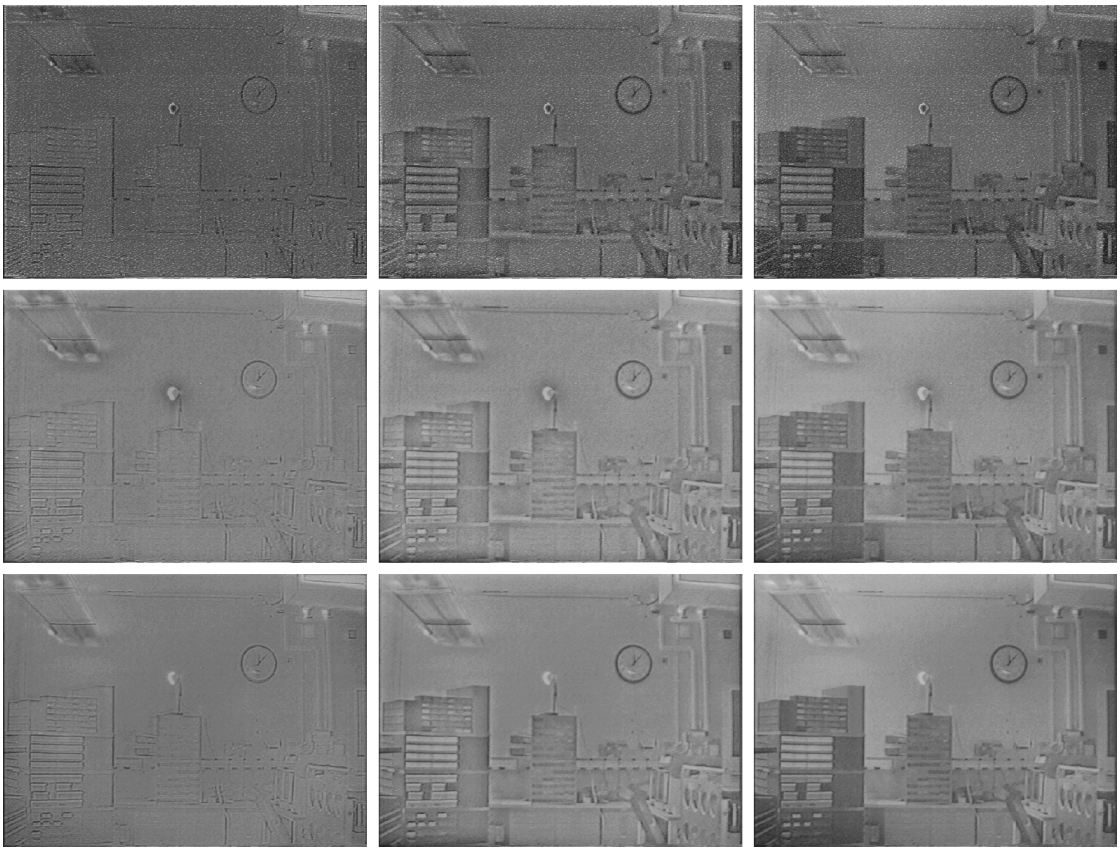

Figure 6. Top row: $(240, 5, 5)$, $(240, 5, 20)$, $(240, 5, 240)$. Middle row: $(240, 20, 5)$, $(240, 20, 20)$, $(240, 20, 240)$. Bottom row: $(240, 240, 5)$, $(240, 240, 20)$, $(240, 240, 240)$.

3. CONCLUSIONS

The task of landing an aircraft in poor visibility conditions is one of the most challenging that a pilot faces. Any technique which allows the pilot to alleviate the problems due to poor visibility will result in safer flights. The two methods we presented in this paper attempt to accomplish this task. Using data from three different sources, long-wave infrared, short-wave infrared, and color, we provide as an end-product, a single fused image that contains more visual information than do any of the data streams individually. The MSR and SSR algorithms are used to process the original data streams, providing enhanced data streams which contain data that is sharper, and has better contrast: i.e., data that provides better visibility.

Two different approaches are presented, both for the same end-goal: better visibility in poor visibility conditions. The techniques differ in some basic ways such as how the different data streams are fused together and which data stream contains more useful features, but each fused data stream has more, and better, information than the original data streams. In addition, the fused data stream allows the data from different sensors to be presented compactly as a single image. This make the pilot's task simpler in that a single image needs to be interpreted rather than three separate ones. If the pilot were required to interpret several data streams, it is extremely possible that one or more of these data streams, which may contain useful information, may be completely ignored, thus incurring errors that could have been avoided.

An issue which has not been touched upon at all in this paper is that of noise enhancement. Noise typically appears as fine detail in an image, and the MSR (and SSR) both will enhance it especially when small scale factors are used. Eventually, the acceptability of noise is observer and application dependent and reduces to this simple determination: Is more information *with* noise better than less information *without* noise? As long

as the noise does not contaminate the information in the signal, the authors prefer more information with noise to less information without. However the choice, as was said earlier, is user and application dependent.

Acknowledgments

Dr. Rahman's work was supported with the NASA cooperative agreement NCC-1-01030.

REFERENCES

1. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
4. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
5. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing* **6**, pp. 965–976, July 1997.
6. B. D. Thompson, Z. Rahman, and S. K. Park, "A multi-scale retinex for improved performance in multi-spectral image classification," in *Visual Information Processing IX*, Proc. SPIE 4041, 2000.
7. B. D. Thompson, Z. Rahman, and S. K. Park, "Retinex preprocessing for improved multi-spectral image classification," in *Visual Information Processing VIII*, Proc. SPIE 3716, 1999.
8. C. L. Tiana, J. R. Kerr, and S. D. Harrah, "Multispectral uncooled infrared enhanced-vision system for flight test," in *Enhanced and Synthetic Vision 2001*, J. G. Verly, ed., pp. 231–236, Proc. SPIE 4363, 2001.
9. L. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**, pp. 325–376, December 1992.
10. P. K. Kaiser and R. M. Boynton, *Human Color Vision*, Optical Society of America, Washington, DC, second ed., 1996.